

SEQUENCE LISTING



(1) GENERAL INFORMATION:

(i) APPLICANT: Kirschbaum, Bernd
Berglund, Erick
Meisterernst, Michael
Polites, Greg

(ii) TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
COMPLEXES FROM TRANSGENIC NON-HUMAN ANIMALS

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: HELLER, EHRMAN, WHITE & MCAULIFFE
(B) STREET: 1666 K Street, N.W., Suite 300
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20006

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 09/849,243
(B) FILING DATE: 07-May-2001

AI (viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Granados, Patricia D.
(B) REGISTRATION NUMBER: 33,683
(C) REFERENCE/DOCKET NUMBER: 38005-0148

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (202)912-2000
(B) TELEFAX: (202)912-2020

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val
1 5 10

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala
1 5

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
 - (A) NAME/KEY: exon
 - (B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGAGCAACCG CCTGCTGGGT GC

22

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
 - (A) NAME/KEY: exon
 - (B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCTGTGTTGC CTGCTGGGAC G

21

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:
 - (A) NAME/KEY: exon

(B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGAGACTGAA GTTAGGCCAG C

21

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 76 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 1..76

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCGGCACCAG GCCGCTGCTG TGATGATGAT GATGATGGCT GCTGCCCATG ACTGCGTAAT

60

GCGGTCATGA CGCTTT

76

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 75 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 1..75

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GAAGGGGGTG GGGGAGGCAA GGGTACATGA GAGCCATTAC GTCGTCTTCC TGAATCCCTT

60

TAGCCGCTTT GCTCG

75

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCCTATGACG TCCCGGATTA CG

22

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..22

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTGGAGTGGT GCCCGCAAG GG

22

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro
1 5 10 15
Arg Gly Cys

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1310 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..1310

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

60 CCATGGGCTA TCCCTATGAC GTCCCGGATT ACGCAGTCAT GGGCAGCAGC CATCATCATC
120 ATCATCACAG CAGCGGCCTG GTGCCGCGCG GCAGCCATAT GGATCAGAAC AACAGCCTGC
180 CACCTTACGC TCAGGGCTTG GCCTCCCCTC AGGGTGCCAT GACTCCCGGA ATCCCTATCT
240 TTAGTCCAAT GATGCCTTAT GGCACTGGAC TGACCCACACA GCCTATTCAG AACACCAATA
300 GTCTGTCTAT TTTGGAAGAG CAACAAAGGC AGCAGCAGCA ACAACAACAG CAGCAGCAGC
360 AGCAGCAGCA GCAGCAACAG CAACAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC
420 AGCAGCAGCA GCAGCAGCAA CAGGCAGTGG CAGCTGCAGC CGTTCAGCAG TCAACGTCCC
480 AGCAGGCAAC ACAGGGAACC TCAGGCCAGG CACCACAGCT CTTCCACTCA CAGACTCTCA
540 CAACTGCACC CTTGCCGGGC ACCACTCCAC TGTATCCCTC CCCCATGACT CCCATGACCC

600 CCATCACTCC TGCCACGCCA GCTTCGGAGA GTTCTGGGAT TGTACCGCAG CTGCAAAATA
 660 TTGTATCCAC AGTGAATCTT GGTGTAAAC TTGACCTAAA GACCATTGCA CTTCGTGCCC
 720 GAAACGCCGA ATATAATCCC AAGCGGTTTG CTGCGGTAAT CATGAGGATA AGAGAGCCAC
 780 GAACCACGGC ACTGATTTTC AGTTCTGGGA AAATGGTGTG CACAGGAGCC AAGAGTGAAG
 840 AACAGTCCAG ACTGGCAGCA AGAAAATATG CTAGAGTTGT ACAGAAGTTG GGTTTTCCAG
 900 CTAAGTTCTT GGACTTCAAG ATTCAGAACA TGGTGGGGAG CTGTGATGTG AAGTTTCCTA
 960 TAAGGTTAGA AGGCCTTGTG CTCACCCACC AACAAATTTAG TAGTTATGAG CCAGAGTTAT
 1020 TTCCTGGTTT AATCTACAGA ATGATCAAAC CCAGAATTGT TCTCCTTATT TTTGTTTCTG
 1080 GAAAAGTTGT ATTAACAGGT GCTAAAGTCA GAGCAGAAAT TTATGAAGCA TTTGAAAACA
 1140 TCTACCCTAT TCTAAAGGGA TTCAGGAAGA CGACGTAATG GCTCTCATGT ACCCTTGCCT
 1200 CCCCCACCCC CTTCTTTTTT TTTTTTTAAA CAAATCAGTT TGTTTTGGTA CCTTTAAATG
 1260 GTGGTGTGTG GAGAAGATGG ATGTTGAGTT GCAGGGTGTG GCACCAGGTG ATGCCCTTCT
 1310 GTAAGTGCCC CTTCCGGCAT CCCGGAATTC CTGCAGCCCA ACGCGGCCGC

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4286 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: exon
- (B) LOCATION: 1..4286

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

60 GAATTCCCCT GCAGGTCAC TAGCGTTGGC CACATAGTAG GTTCTCAAAT ACTTGTTAAT
 120 AAATAAGTTT GTTCGAGAAG CTGGGCAATG ATATTCTACA GCTGGAAGAA GAAACATAAT
 180 GATCTAGTAA TTAGCTCAAT TAAAAATAAA CGTTCTTCTT TCCTCAGAGG AGCATTTCCT
 240 AAGGCCTGCC TTGATAGCCA TCCAAAAAGG CCAAGCTCAT CCAATCTTGC CCTAGATTTA
 300 TGCTAAAATG CAGTTACAAT CGATAGGATG ACAGAAAACG ACAGCACTTA TTAAATATA
 360 ATAGGCACTT ATTTAAATAG GAGAAGCTGT GACTTCATAG CAAGTGTTGG GGTTAGGAAA
 420 CTGGGTGGAT AAAGTTGCTG ATGCTGTAGA TCTTAGCCTC TACATGAGAT CATGTGGAAA
 480 ATCTGAAAGC ATTTTAGGTT CCTTATGTTT GCAATCAAAT AACTGTACAC CTTTAAATTT
 540 AAAAAGTACC ATGAGGCACA CACACACACT CGCAGGAACT TTTTGGCGTA ACAAAGTAG
 600 AATTAGATCT AAAAGCTAAC TGTAGGACTG AGTCTATTCT AAAGTAAAG CCTGGACATC
 660 TGGAGTACCA GGGGGAGATG ACGTGTTACG GGCTTCATA AAAGCAGCTG GCTTTGAATG
 720 GAAGGAGCCA AGAGGCCAGC ACAGGAGCGG ATTCGTCGCT TTCACGGCCA TCGAGCCGAA
 780 CCTCTCGCAA GTCCGTGAGC CGTTAAGGAG GCCCCAGTC CCGACCCTTC GCCCCAAGCC
 840 CCTCGGGGTC CCCGGGCTG GTACTCCTTG CCACACGGGA GGGGCGCGGA AGCCGGGGCG
 900 GAGGAGGAGC CAACCCCGGG CTGGGCTGAG ACCCGCAGAG GAAGACGCTC TAGGGATTG
 960 TCCCGGACTA GCGAGATGGC AAGGCTGAGG ACGGGAGGCT GATTGAGAGG CGAAGGTACA
 1020 CCCTAATCTC AATACAACCT TTGGAGCTAA GCCAGCAATG GTAGAGGGAA GATTCTGCAC
 1080 GTCCCTTCCA GGCGGCCTCC CCGTCACCAC CCCCCCAAC CCGCCCCGAC CGGAGCTGAG
 1140 AGTAATTCAT ACAAAGGAC TCGCCCCTGC CTTGGGGAAT CCCAGGGACC GTCGTAAAC

1200 TCCCACTAAC GTAGAACCCA GAGATCGCTG CGTTCCCGCC CCCTCACCCG CCCGCTCTCG
 1260 TCATCACTGA GGTGGAGAAG AGCATGCGTG AGGCTCCGGT GCCCGTCAGT GGGCAGAGCG
 1320 CACATCGCCC ACAGTCCCCG AGAAGTTGGG GGGAGGGGTC GGCAATTGAA CCGGTGCCTA
 1380 GAGAAGGTGG CGCGGGGTAA ACTGGGAAAAG TGATGTCGTG TACTGGCTCC GCCTTTTTTC
 1440 CGAGGGTGGG GGAGAACCGT ATATAAGTGC AGTAGTCGCC GTGAACGTTC TTTTTCGCAA
 1500 CGGGTTTGCC GCCAGAACAC AGGTAAGTGC CGTGTGTGGT TCCCGCGGGC CTGGCCTCTT
 1560 TACGGGTAT GGCCTTGCG TGCCTTGAAT TACTTCCACG CCCCTGGCTG CAGTACGTGA
 1620 TTCTTGATCC CGAGCTTCGG GTTGAAGTG GGTGGGAGAG TTCGAGGCCT TCGCTTAAG
 1680 GAGCCCCTTC GCCTCGTGCT TGAGTTGAGG CCTGGCCTGG GCGCTGGGGC CGCCGCGTGC
 1740 GAATCTGGTG GCACCTTCGC GCCTGTCTCG CTGCTTTCGA TAAGTCTCTA GCCATTTAAA
 1800 ATTTTGTATG ACCTGCTGCG ACGCTTTTTT TCTGGCAAGA TAGTCTTGTA AATGCGGGCC
 1860 AAGATCTGCA CACTGGTATT TCGGTTTTTG GGGCCGCGGG CGGCGACGGG GCCCGTGCCT
 1920 CCCAGCGCAC ATGTTCTGGCG AGGCGGGGCC TCGAGCGCG GCCACCGAGA ATCGGACGGG
 1980 GGTAGTCTCA AGCTGGCCGG CCTGCTCTGG TGCCTGGCCT CGCGCCGCCG TGTATCGCCC
 2040 CGCCCTGGGC GGCAAGGCTG GCGCGGTCGG CACCAGTTGC GTGAGCGGAA AGATGGCCGC
 2100 TTCCCGGCCC TGCTGCAGGG AGCTCAAAAT GGAGGACGCG GCGCTCGGGA GAGCGGGCGG
 2160 GTGAGTCACC CACACAAAGG AAAAGGGCCT TTCCGTCCTC AGCCGTCGCT TCATGTGACT
 2220 CCACGGAGTA CCGGGCGCCG TCCAGGCACC TCGATTAGTT CTCGAGCTTT TGGAGTACGT
 2280 CGTCTTTAGG TTGGGGGGAG GGGTTTTATG CGATGGAGTT TCCCACACT GAGTGGGTGG

2340 AGACTGAAGT TAGGCCAGCT TGGCACTTGA TGTAATTCTC CTTGGAATTT GCCCTTTTTG
 2400 AGTTTGGATC TTGGTTCATT CTCAAGCCTC AGACAGTGGT TCAAAGTTTT TTTCTTCCAT
 2460 TTCAGGTGTC GTGAGGAATT GCCCAGGGGA TCCATGGGCT ATCCCTATGA CGTCCCGGAT
 2520 TACGCAGTCA TGGGCAGCAG CCATCATCAT CATCATCACA GCAGCGGCCT GGTGCCGCGC
 2580 GGCAGCCATA TGGATCAGAA CAACAGCCTG CCACCTTACG CTCAGGGCTT GGCCTCCCCCT
 2640 CAGGGTGCCA TGA CTCCCGG AATCCCTATC TTTAGTCCAA TGATGCCTTA TGGCACTGGA
 2700 CTGACCCAC AGCCTATTCA GAACACCAAT AGTCTGTCTA TTTTGGAAGA GCAACAAAGG
 2760 CAGCAGCAGC AACAACAACA GCAGCAGCAG CAGCAGCAGC AGCAGCAACA GCAACAGCAG
 2820 CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGGCAGTG
 2880 GCAGCTGCAG CCGTTCAGCA GTCAACGTCC CAGCAGGCAA CACAGGGAAC CTCAGGCCAG
 2940 GCACCACAGC TCTTCCACTC ACAGACTCTC ACAACTGCAC CCTTGCCGGG CACCACTCCA
 3000 CTGTATCCCT CCCCCATGAC TCCCATGACC CCCATCACTC CTGCCACGCC AGCTTCGGAG
 3060 AGTTCTGGGA TTGTACCGCA GCTGCAAAAT ATTGTATCCA CAGTGAATCT TGGTTGTAAA
 3120 CTTGACCTAA AGACCATTGC ACTTCGTGCC CGAAACGCCG AATATAATCC CAAGCGGTTT
 3180 GCTGCGGTAA TCATGAGGAT AAGAGAGCCA CGAACCACGG CACTGATTTT CAGTTCTGGG
 3240 AAAATGGTGT GCACAGGAGC CAAGAGTGAA GAACAGTCCA GACTGGCAGC AAGAAAATAT
 3300 GCTAGAGTTG TACAGAAGTT GGGTTTTCCA GCTAAGTTCT TGGACTTCAA GATTCAGAAC
 3360 ATGGTGGGGA GCTGTGATGT GAAGTTTCCT ATAAGGTTAG AAGGCCTTGT GCTCACCAC
 3420 CAACAATTTA GTAGTTATGA GCCAGAGTTA TTTCTGGTT TAATCTACAG AATGATCAAA

3480 CCCAGAATTG TTCTCCTTAT TTTTGTCTTCT GGAAAAGTTG TATTAACAGG TGCTAAAGTC
 3540 AGAGCAGAAA TTTATGAAGC ATTTGAAAAC ATCTACCCTA TTCTAAAGGG ATTCAGGAAG
 3600 ACGACGTAAT GGCTCTCATG TACCCTTGCC TCCCCACCC CCTTCTTTTT TTTTTTTTAA
 3660 ACAAATCAGT TTGTTTTGGT ACCTTTAAAT GGTGGTGTTG TGAGAAGATG GATGTTGAGT
 3720 TGCAGGGTGT GGCACCAGGT GATGCCCTTC TGTAAGTGCC CCTTCCGGCA TCCCGGATAT
 3780 CCTGCAGCCC AACACGGCCG CTCGAGCATG CATCTAGAGA ACGTCACGGC CGCGATCCCC
 3840 CTGTGCCTTC TAGTTGCCAG CCATCTGGTT GTTTGCCCCCT CCCCCGTGCC TTCCTTGACC
 3900 CTGGAAGGTG CCACTCCCAC TGTCTTTTCC TAATAAAATG AGGAAATTGC ATCGCATTGT
 3960 CTGAGTAGGT GTCATTCTAT TCTGGGGGGT GGGGTGGGGC AGGACAGCAA GGGGGAGGAT
 4020 TGGGAAGACA ATAGCAGGCA TGCTGGGGAT GCGGTGGGCT CTATGGGTAC CCAGGTGCTG
 4080 AAGAATTGAC CCGGTTCTCT CTGGGCCAGA AAGAAGCAGG CACATCCCCT TCTCTGTGAC
 4140 ACACCCTGTC CACGCCCCCTG GTTCTTAGTT CCAGCCCCAC TCATAGGACA CTCAACTTGG
 4200 AGCGGTCTCT CCCTCCCTCA TCAGCCCACC AAACCAAACC TAGCCTCCA GAGTGGGAAG
 4260 AAATTAAAGC AAGAAGGCTA TTAAGTGCA GGGGAGAGAA AATGCCTCCA ACATGTGAGG
 4286 AAGTAATGAT AGAAATCATA GAATTC

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3263 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: exon

(B) LOCATION: 1..3263

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

60 ATCGATAAGC TGAGATCCGG CTAGAAACTG CTGAGGGCTG GACCGCATCT GGGGACCATC
120 TGTTCCTGGC CCTGAGCGGG GCAGGAACTG CTTACCGCAG ATATCCTGTT TGCCCCAATT
180 CAGCTGTTCC ATCTGTTCTT GGCCCTGAGC GGGGCAGGAA CTGCTTACCA CAGATATCCT
240 GTTTGGCCCA TATTCAGCTG TCTCTCTGTT CCTGACCTTG ATCTGAACTT CTCTATTCTC
300 AGTTATGTAT TTTTCCCATG CCTTGCAAAA TGGCGTTACT TAAGCTAGCT TGCCAAACCT
360 ACGGCTGGGG TCTTTCACGT TTATATCTAT GAGGGGAAGG ACCCAGAGTG GGAAGCTGG
420 GATCTTGGA ACACGCTTCT CTACATGGCA TTGTCTGCAC GGTGGAGTCC GGATCTGAGC
480 TTGGCTTGGT TTTTAAAACC AGCCTGGAGT AGAGCAGATG GGTTAAGGTG AGTGACCCCT
540 CAGCCCTGGA CATTCTTAGA TGAGCCCCCT CAGGAGTAGA GAATAATGTT GAGATGAGTT
600 CTGTTGGCTA AAATAATCAA GGCTAGTCTT TATAAACTG TCTCCTCTTC TCCTAGCTTC
660 GATCCAGAGA GAGACCTGGG CGGAGCTGGT CGCTGCTCAG GAACTCCAGG AAAGGAGAAG
720 CTGAGGTTAC CACGCTGCGA ATGGGTTTAC GGAGATAGCT GGCTTTCCGG GGTGAGTTCT
780 CGTAAACTCC AGAGCAGCGA TAGGCCGTAA TATCGGGGAA AGCACTATAG GGACATGATG
840 TTCCACACGT CACATGGGTC GTCCTATCCG AGCCAGTCGT GCCAAAGGGG CGGTCCCGCT
900 GTGCACACTG GCGCTCCAGG GAGCTCTGCA CTCCGCCCGA AAAGTGCGCT CGGCTCTGCC
960 AGGACGCGGG GCGCGTGA CTGCGTGGGC TGGAGCAACC GCCTGCTGGG TGCAAACCCT

1020 TTGCGCCCGG ACTCGTCCAA CGACTATAAA GAGGGCAGGC TGTCTCTAA GCGTCACCAC
 1080 GACTTCAACG TCCTGAGTAC CTTCTCCTCA CTTACTCCGT AGCTCCAGCT TCACCAGATC
 1140 CTCGAGAACG TCTCCCATGG GCTATCCCTA TGACGTCCCG GATTACGCAG TCATGGGCAG
 1200 CAGCCATCAT CATCATCATC ACAGCAGCGG CCTGGTGCCG CGCGGCAGCC ATATGGATCA
 1260 GAACAACAGC CTGCCACCTT ACGCTCAGGG CTTGGCCTCC CCTCAGGGTG CCATGACTCC
 1320 CGGAATCCCT ATCTTTAGTC CAATGATGCC TTATGGCACT GGACTGACCC CACAGCCTAT
 1380 TCAGAACACC AATAGTCTGT CTATTTTGGA AGAGCAACAA AGGCAGCAGC AGCAACAACA
 1440 ACAGCAGCAG CAGCAGCAGC AGCAGCAGCA ACAGCAACAG CAGCAGCAGC AGCAGCAGCA
 1500 GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA GCAACAGGCA GTGGCAGCTG CAGCCGTTCA
 1560 GCAGTCAACG TCCCAGCAGG CAACACAGGG AACCTCAGGC CAGGCACCAC AGCTCTTCCA
 1620 CTCACAGACT CTCACAAGT CACCCTTGCC GGGCACCCT CCACTGTATC CCTCCCCCAT
 1680 GACTCCCATG ACCCCCATCA CTCCTGCCAC GCCAGCTTCG GAGAGTTCTG GGATTGTACC
 1740 GCAGCTGCAA AATATTGTAT CCACAGTGAA TCTTGTTGT AAATTGACC TAAAGACCAT
 1800 TGCACTTCGT GCCCGAAACG CCGAATATAA TCCAAGCGG TTGCTGCGG TAATCATGAG
 1860 GATAAGAGAG CCACGAACCA CGGCACTGAT TTTAGTTCT GGGAAAATGG TGTGCACAGG
 1920 AGCCAAGAGT GAAGAACAGT CCAGACTGGC AGCAAGAAAA TATGCTAGAG TTGTACAGAA
 1980 GTTGGGTTTT CCAGCTAAGT TCTTGACTT CAAGATTCAG AACATGGTGG GGAGCTGTGA
 2040 TGTGAAGTTT CCTATAAGGT TAGAAGGCCT TGTGCTCACC CACCAACAAT TTAGTAGTTA
 2100 TGAGCCAGAG TTATTTCTTG GTTTAATCTA CAGAATGATC AAACCCAGAA TTGTTCTCCT

2160 TATTTTTGTT TCTGGAAAAG TTGTATTAAC AGGTGCTAAA GTCAGAGCAG AAATTTATGA
 2220 AGCATTTGAA AACATCTACC CTATTCTAAA GGGATTCAGG AAGACGACGT AATGGCTCTC
 2280 ATGTACCCTT GCCTCCCCCA CCCCTTCTT TTTTTTTTTT TAAACAAATC AGTTTGTTTT
 2340 GGTACCTTTA AATGGTGGTG TTGTGAGAAG ATGGATGTTG AGTTGCAGGG TGTGGCACCA
 2400 GGTGATGCCC TTCTGTAAGT GCCCCCTCCG GCATCCCGGA ATTCCTGCAG CCCAACGCGG
 2460 CCGCTTCGAG GGATCTTTGT GAAGGAACCT TACTTCTGTG GTGTGACATA ATTGGACAAA
 2520 CTACCTACAG AGATTTAAAG CTCTAAGGTA AATATAAAAT TTTTAAGTGT ATAATGTGTT
 2580 AAACACTGA TTCTAATTGT TTGTGTATTT TAGATTCCAA CCTATGGAAC TGATGAATGG
 2640 GAGCAGTGGT GGAATGCCTT TAATGAGGAA AACCTGTTTT GCTCAGAAGA AATGCCATCT
 2700 AGTGATGATG AGGCTACTGC TGACTCTCAA CATTCTACTC CTCCAAAAAA GAAGAGAAAAG
 2760 GTAGAAGACC CCAAGGACTT TCCTTCAGAA TTGCTAAGTT TTTTGAGTCA TGCTGTGTTT
 2820 AGTAATAGAA CTCTTGCTTG CTTTGCTATT TACACCACAA AGGAAAAAGC TGCACTGCTA
 2880 TACAAGAAAA TTATGGAAAA ATATTCTGTA ACCTTTATAA GTAGGCATAA CAGTTATAAT
 2940 CATAACATAC TGTTTTTTCT TACTCCACAC AGGCATAGAG TGTCTGCTAT TAATAACTAT
 3000 GCTCAAAAAT TGTGTACCTT TAGCTTTTTA ATTTGTAAAG GGGTTAATAA GGAATATTTG
 3060 ATGTATAGTG CCTTGACTAG AGATCATAAT CAGCCATACC ACATTTGTAG AGGTTTTACT
 3120 TGCTTTAAAA AACCTCCCAC ACCTCCCCCT GAACCTGAAA CATAAAATGA ATGCAATTGT
 3180 TGTTGTTAAC TTGTTTATTG CAGCTTATAA TGTTTACAAA TAAAGCAATA GCATCACAAA
 3240 TTTCACAAAT AAAGCATTTT TTTCACCTGCA TTCTAGTTGT GGTTTGTCCA AACTCATCAA

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 371 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```

Met Gly Tyr Pro Tyr Asp Val Pro Asp Tyr Ala Val Met Gly Ser Ser
1           5           10           15

His His His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser His
          20           25           30

Met Asp Gln Asn Asn Ser Leu Pro Pro Tyr Ala Gln Gly Leu Ala Ser
          35           40           45

Pro Gln Gly Ala Met Thr Pro Gly Ile Pro Ile Phe Ser Pro Met Met
          50           55           60

Pro Tyr Gly Thr Gly Leu Thr Pro Gln Pro Ile Gln Asn Thr Asn Ser
65           70           75           80

Leu Ser Ile Leu Glu Glu Gln Gln Arg Gln Gln Gln Gln Gln Gln Gln
          85           90           95

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
          100          105          110

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Ala
          115          120          125

Val Ala Ala Ala Ala Val Gln Gln Ser Thr Ser Gln Gln Ala Thr Gln
          130          135          140

Gly Thr Ser Gly Gln Ala Pro Gln Leu Phe His Ser Gln Thr Leu Thr
145          150          155          160

Thr Ala Pro Leu Pro Gly Thr Thr Pro Leu Tyr Pro Ser Pro Met Thr
          165          170          175

Pro Met Thr Pro Ile Thr Pro Ala Thr Pro Ala Ser Glu Ser Ser Gly
          180          185          190

Ile Val Pro Gln Leu Gln Asn Ile Val Ser Thr Val Asn Leu Gly Cys
          195          200          205

Lys Leu Asp Leu Lys Thr Ile Ala Leu Arg Ala Arg Asn Ala Glu Tyr

```

210 215 220
 Asn Pro Lys Arg Phe Ala Ala Val Ile Met Arg Ile Arg Glu Pro Arg
 225 230 235 240
 Thr Thr Ala Leu Ile Phe Ser Ser Gly Lys Met Val Cys Thr Gly Ala
 245 250
 Lys Ser Glu Glu Gln Ser Arg Leu Ala Ala Arg Lys Tyr Ala Arg Val
 260 265 270
 Val Gln Lys Leu Gly Phe Pro Ala Lys Phe Leu Asp Phe Lys Ile Gln
 275 280 285
 Asn Met Val Gly Ser Cys Asp Val Lys Phe Pro Ile Arg Leu Glu Gly
 290 295 300

 Leu Val Leu Thr His Gln Gln Phe Ser Ser Tyr Glu Pro Glu Leu Phe
 305 310 315 320
 Pro Gly Leu Ile Tyr Arg Met Ile Lys Pro Arg Ile Val Leu Leu Ile
 325 330 335
 Phe Val Ser Gly Lys Val Val Leu Thr Gly Ala Lys Val Arg Ala Glu
 340 345 350
 Ile Tyr Glu Ala Phe Glu Asn Ile Tyr Pro Ile Leu Lys Gly Phe Arg
 355 360 365
 Lys Thr Thr
 370

A1
 (end)

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Met Gly Ser Ser His His His His His Ser Ser Gly Leu Val Pro
 1 5 10 15
 Arg Gly